Meta-Mass Shift Chemical (MeMSChem) profiling of metabolomes from coral reefs

Hartmann, Aaron; Petras, Daniel; Quinn, Robert; Protsyuk, Ivan; Archer, Frederick; Ransome, Emma; Williams, Gareth; Bailey, Barbara; Vermeij, Mark; Alexandrov, Theodore; Dorrestein, Pieter; Rohwer, Forest

Proceedings of the National Academy of Sciences of the United States of America

DOI:
10.1073/pnas.1710248114

Published: 31/10/2017

Peer reviewed version

Meta-Mass Shift Chemical (MeMSChem) profiling of metabolomes from coral reefs

Aaron C. Hartmann[1,2,*], Daniel Petras[3], Robert A. Quinn[3], Ivan Protsyuk[4], Frederick I. Archer[5], Emma J. Ransome[2], Gareth J. Williams[6], Barbara A. Bailey[7], Mark J. A. Vermeij[8,9], Theodore Alexandrov[3,4], Pieter C. Dorrestein[3], Forest L. Rohwer[1]

Affiliations

[1] Department of Biology, San Diego State University, San Diego, CA
[2] National Museum of Natural History, Smithsonian Institution, Washington, D.C.
[3] Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Science, University of California San Diego, La Jolla, CA
[4] Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany
[5] National Oceanic and Atmospheric Administration, Southwest Fisheries Science Center, La Jolla, CA
[6] School of Ocean Sciences, Bangor University, LL59 5AB, UK
[7] Department of Mathematics and Statistics, San Diego State University, San Diego, CA
[8] Carmabi Foundation, Piscaderabaai, Willemstad, Curaçao
[9] Aquatic Microbiology, Institute for Biodiversity and Ecosystem Dynamics (IBED) University of Amsterdam, Amsterdam, the Netherlands

*__Corresponding author__:
Department of Invertebrate Zoology
MRC 163 PO BOX 37012
National Museum of Natural History
10th Street and Constitution Avenue NW
Washington DC 20013-7012
aaron.hartmann@gmail.com
+1.802.279.8109

Short title: Meta-Mass Shift Chemical profiling

Classification: Biological Sciences/Biochemistry

Keywords: untargeted metabolomics, molecular networking, small molecules, coral reefs

Abstract

Untargeted metabolomics of environmental samples routinely detects thousands of small molecules, the vast majority of which cannot be identified. Meta-Mass Shift Chemical (MeMSChem) profiling was developed to identify mass differences between related molecules using metabolomic networks. This approach illuminates metabolome-wide relationships between molecules and the putative chemical groups that differentiate them (e.g., H_2, CH_2, $COCH_2$). MeMSChem was used to analyze a publicly-available metabolomic dataset of coral, algal, and fungal mat holobionts (i.e., the host and its associated microbes and viruses) sampled from some of Earth's most remote and pristine coral reefs. Each type of holobiont had unique mass shift profiles, even when the analysis was restricted to parent molecules found in all samples. This result suggests that holobionts modify the same molecules in different ways and offers insights into the generation of molecular diversity. Three genera of stony corals had distinct patterns of molecular relatedness despite their high degree of taxonomic relatedness. MeMSChem profiles also partially differentiated between individuals, suggesting that every coral reef holobiont is a potential source of novel chemical diversity.

Significance Statement

Coral reef taxa produce a diverse array of molecules, some of which are important pharmaceuticals. To better understand how molecular diversity is

64 generated on coral reefs, metabolomes were analyzed using a novel approach
65 called Meta-Mass Shift Chemical (MeMSChem) profiling. MeMSChem uses the
66 mass differences between molecules in networks to determine how molecules
67 are related. Interestingly, the same molecules gain and lose chemical groups in
68 different ways depending on the taxa it came from, offering a partial explanation
69 for high molecular diversity on coral reefs.

\body

Untargeted tandem mass spectrometry is a powerful tool for wide-scale analysis of small molecules. The resulting metabolomes are potential treasure troves of new molecules and chemistries, but a major problem in realizing this potential is that most detected molecules cannot be identified (1-5). One possible solution is to use spectral fragmentation similarity to identify relatives of known molecules to generate novel annotations (6-8). These approaches have rapidly expanded reference databases, but they remain inherently limited by the number of known molecules. Therefore, there is a need for analyses that do not rely upon molecular reference libraries (9).

The online platform Global Natural Products Social Molecular Networking (GNPS; 5) uses spectral fragmentation patterns to compare tens of thousands of molecular features and create networks of structurally-similar molecules. Here we expand the analysis of GNPS networks to identify chemical differences between related molecules (Figure 1). This approach is called Meta-Mass Shift Chemical (MeMSChem) profiling and it uses the mass differences (or mass shifts) between related molecules to identify and annotate known chemical groups such as H_2, CH_2, $COCH_2$, etc. Annotating molecules based on their mass shifts facilitates correlations between metabolomics, biochemistry and genomics, which could help pinpoint sites of molecular modifications resulting from known and unknown enzymatic activities.

Coral reefs are noted sources of novel, commercially-useful compounds (10). Reef holobionts (e.g., corals, sponges, and algae with their associated viral and microbial communities; 11) have unique metabolomes, with a high degree of within-holobiont similarity (12, 13). The positive relationship between taxonomic and molecular diversity is evident at the ecosystem level, but mechanisms explaining how high molecular diversity is generated remain missing. To address this question, MeMSChem was applied to an existing dataset (12) comprised of seven coral reef holobiont types collected in the Line Islands, which are some of the most remote and pristine coral reefs in the world (14, 15). MeMSChem profiling showed that molecular mass shift patterns differ significantly between holobionts, offering new insights into why high molecular diversity is found on coral reefs.

Results

Identifying redundant mass shifts in metabolomes in coral reef holobionts: The dataset used as the basis for creating MeMSChem was previously published in Quinn et al. 2016 (12) and can be found on the Mass spectrometry Interactive Virtual Environment (MassIVE) at https://massive.ucsd.edu/ with the accession number: MSV000078598. This dataset was derived from an LC-MS/MS analysis of three genera of scleractinian coral (*Montipora* spp., *Pocillopora* spp., *Porites* spp.), two coralline algae (crustose coralline algae [CCA] and *Halimeda* sp.), two non-calcifying algae (macroalgae and turf algae) and a fungal mat.

The online platform Global Natural Products Social Molecular Networking (GNPS, gnps.ucsd.edu; Figure S1; 5) was used to cluster identical MS/MS spectra into nodes and identify the degree to which each node was structurally similar (i.e., related) to other nodes (henceforth referred to as molecular features) based on a cosine score of spectral similarity. All pairs of molecular features receiving a cosine score above a threshold of 0.6 were considered to be related and connected in the network (see SI Materials and Methods for more details regarding the cosine score threshold). Each mass shift between network connections was then mined for multiple (i.e., redundant) occurrences (Figure 1B). When five or more molecular feature pairs differed by a mass of *m/z* ± 0.001, the mass shift was counted. All molecular features comprising the pairs with this mass shift were assigned to a bin (Figure 1C-E, see the SI Materials and Methods for more details).

MeMSChem profiling identified 62 mass shifts that passed the filter of *m/z* ± 0.001 and greater than 5 occurrences (Table S1). Among these mass shifts, 10 were annotated to known adducts and artifacts and were removed prior to further analyses (Table S1 and S2). The remaining mass shifts were annotated to known chemical groups involving hydrogen, carbon, and oxygen where possible, leading to the annotation of 13 of the 62 mass shifts identified (Table S1). This represents a conservative list of annotations and the additional mass shifts identified here may be annotatable in future investigations.

Mass shifts of 0 were abundant in the networks and may represent isomers. These mass shifts were removed due to the likelihood that two isomers were merged into a single molecular feature or that the same molecular feature was split into two molecules during networking, due to the high degree or spectral similarity or differences in the number of observable fragments, respectively. An approach using retention time differences or chiral separation columns should be employed to separate isomers in future applications of MeMSChem.

Quantifying mass shifts in holobionts: MS/MS-based molecular features associated with the redundant mass shifts were quantified from the MS scan of the parent molecule using the Optimus software (https://github.com/MolecularCartography/Optimus; Figure 1C). A molecular feature filter was applied to remove features that were not detected in all samples. Consequently, only the features present in all samples were quantified. This filter allowed us to determine whether holobionts exhibit different mass shifts associated with the same molecules (c.f., different mass shifts associated with molecules that are only found that holobiont; Figure 1E-F).

Three forms of metabolome-wide data were generated for each sample (Figure 1A-C). First, all instances where a redundant mass shift was observed in the network was tabulated for each sample. These 'counts' data provided a metric of the commonness and rarity of each mass shift in each sample (Figure 1D). Second, the abundance of every molecular feature was summed by mass

156 shift regardless of whether that feature was the higher or lower mass feature in a
157 network pair. These 'summed abundance' data provided a metric for the overall
158 occurrence of each mass shift throughout each sample (Figure 1E; see the SI
159 Materials and Methods for equations). Third, for each network pair, the difference
160 in abundance between the more and less massive feature was calculated, then
161 these values were summed by mass shift for each sample (Figure 1F, see the SI
162 Materials and Methods for equations). These 'difference in abundance' data
163 reflected whether, metabolome-wide, molecules were more likely to gain or lose
164 a given mass, potentially reflecting active interconversion or branching of largely
165 shared biosynthetic pathways. All resultant data are provided in the Supporting
166 Information 'Processed Data' file. Among the redundant mass shifts, seven of the
167 ten most redundant mass shifts were putatively annotated to known chemical
168 groups, constituting nearly 50% of the network pairs isolated from the networks.
169 These mass shifts included *m/z* 2.016, 14.016, 28.032, 56.064, 26.016, 18.010,
170 12.000, which were putatively annotated as H_2, CH_2, C_2H_4, C_4H_8, C_2H_2, H_2O, and
171 C, respectively.

172 *Examining known mass shifts associated with library-identified molecular*
173 *features:* Instances in which known mass shifts were associated with identified
174 molecules provided conformational evidence that mass shifts were correctly
175 annotated. Four examples are highlighted in Figure 2B: 1) A feature identified as
176 phenatro-furanone with a mass shift of *m/z* 18.014 (H_2O; Figure 2B Example 1;
177 Figure S2). 2) A subnetwork with three forms of Lyso-PAF and related

178 compounds (Figure 2B Example 2; Figure S3). The identification of one
179 molecular feature, Lyso-PAF C-16, in these sample was previously confirmed
180 using a reference standard by Quinn et al. 2016 (12). This subnetwork is
181 particularly informative because the three identified compounds were networked
182 to one another, showing that the mass shifts truly correspond to a de-saturation
183 and elongation of a fatty acid chain, *m/z* 2.018 (H_2) and *m/z* 28.032 (C_2H_4). 3) A
184 subnetwork of ceramide-related compounds (Figure 2B Example 3; Figure S4)
185 with mass shifts of *m/z* 2.015 (H_2), *m/z* 14.015 (CH_2), and 165.057 ($C_6H_{10}O_5$,
186 glycosylation). A coral-associated ceramide was recently identified (16) with one
187 additional desaturation relative to the ceramide identified here, and this newly-
188 identified ceramide has an extremely similar mass (*m/z* 536.504) to the unknown
189 feature (*m/z* 536.508) networked to the ceramide here. The newly-identified
190 ceramide also differs in mass from the identified ceramide by *m/z* 2.015,
191 consistent with one fewer saturation. 4) A subnetwork of three unidentified
192 molecules with mass shifts of *m/z* 28.032 (C_2H_4), *m/z* 28.033 (C_2H_4), and *m/z*
193 56.065 (C_4H_8; Figure S5).
194 *Differences in mass shift profiles between types of holobionts:* To
195 determine how well MeMSChem profiling resolved each holobiont type, Random
196 Forests classification (17) was used to generate an out-of-bag error (henceforth
197 referred to as 'model error'), which reflects the extent to which the model
198 correctly categorized every sample (i.e., whether *Halimeda* sp. samples were
199 correctly placed into the model's *Halimeda* group). Random Forests offers

200 exceptional classification performance and is robust to non-normally distributed
201 data and correlated predictors (18), both of which were present in this dataset
202 (see the Supporting Information 'Processed Data' file).
203 The usefulness of re-categorizing molecules by their mass shifts was first
204 evaluated based on the number of times that each mass shift was observed
205 (counts data described above). The model error of the Random Forests
206 classifying holobiont types using the counts data was 0.44, which indicates that
207 44% of the time samples were assigned to the incorrect holobiont type. The
208 resolution gained from the observed counts data (i.e., actual data) was compared
209 to that from 1000 permutations of the data in which pairs were randomly binned
210 and counted while keeping the original proportions consistent. The observed
211 counts data outperformed 95% of the randomly-generated datasets, suggesting
212 that the counts of redundant mass shifts aided in differentiating between
213 holobiont types despite the relatively high model error (Figure 3A).
214 Molecular abundance data were then incorporated into the analysis and
215 compared against the holobiont resolution gained from the counts data. When
216 the summed abundances of each mass shift among molecules present in all
217 holobionts were considered, the model error from the abundance data was 0.36
218 (Figure 3B). Therefore, incorporating feature abundance data improved the
219 accuracy of the model by 8% when resolving between holobiont types. The value
220 of summing feature abundances by mass shift was also tested by comparing its
221 accuracy to the models of 1000 permutations of the data in which network pairs

were randomly binned and summed while keeping the original proportions consistent (as was done for the counts data above). Among only the molecular features present in all holobionts, summing of feature abundances by mass shift resolved holobiont types better than 90% of the datasets generated from random summing of feature abundances (Figure 3B). Thus, binning abundance data by redundant mass shifts categorizes molecules in a non-random manner. Molecular abundances binned by mass shifts also reflected differences among holobiont types better than when holobionts were compared with data that lacks any feature abundance information (i.e., counts of the number of mass shifts).

To determine whether mass shifts may reflect active sites of molecular interconversions, as would be expected if a molecular modification had occurred, the summed abundances were compared to the differences in abundance between molecular pairs by mass shift. This is akin to one molecule being the reactant and the other the product. The model error of the difference in abundance data was 0.34, demonstrating that organizing the data by the differences in abundances slightly outperformed the summed abundance data (model error = 34% and 36%, respectively; Figure 3C). When compared to 1000 random permutations of the actual data, the difference in abundance data outperformed 86% of the random models.

Classification was further improved by incorporating the full molecular dataset, and thus the molecules that were present in all holobionts and the molecules that were only found in one or a few holobionts. When these

molecules were included, the model error was 0.02. This reflects a 32% decrease in the model error relative to when only molecules found in all holobionts were considered and was nearly perfect in assigning samples to their correct holobiont type. The real data outperformed 92% of randomly generated datasets (Figure 3D and summarized in Figure 3E). These results suggest that the highest level of holobiont resolution is achieved when: 1) molecular features were binned by the redundant mass shifts they exhibit, 2) molecular abundances were included as the difference in abundance between molecules in a network pair, and 3) molecules/pairs that are only found in certain holobionts were included in addition to those molecules present in all holobionts.

Mass shifts that best distinguish each holobiont type: Among the molecular features present in all holobionts, coral genera were best differentiated from one another by mass shifts corresponding to two carbons that were either saturated (*m/z* 28.032, C_2H_4 or 2*CH_2) or unsaturated (*m/z* 26.016 C_2H_2; $p < 0.01$ for both; Figure 4A). The three coral genera exhibited distinct patterns between these two mass shifts: molecular features of *Montipora* exhibited the addition of C_2H_4 and loss of C_2H_2, while *Pocillopora* exhibited the opposite pattern. *Porites*-associated molecules did not gain or lose either mass shift. Putative CH_2 and CH_2OOH mass shifts best differentiated the non-coral holobionts ($p < 0.01$ for both; Figure 4B). *Halimeda* features predominantly gained CH_2, as did turf algae, the fungal mat, and all of the corals, though to a lesser degree than *Halimeda*. Additions of CH_2OOH were unique to *Halimeda*. Corals were best differentiated

from non-corals based on larger losses of CO and H_2, the latter suggesting a dehydrogenated state perhaps due to higher concentrations of unsaturated lipids.

Discussion

MeMSChem profiling provides an approach to identify mass shifts between related molecules and assign them to known chemical groups in complex metabolomes. Seven coral reef holobiont types were well resolved by MeMSChem profiling. Even among molecular features detected in all holobionts, mass shift profiles differed among holobiont types, suggesting that each type of holobiont is modifying the same molecules in different ways. The chemical differences between holobionts was much more apparent when all molecules were considered (i.e., molecules only produced by certain holobionts were also incorporated), suggesting that disparate mass shift patterns play a role in generating molecular diversity in this ecosystem. Shifts in the abundance of molecules exhibiting each mass shift better resolved holobiont types than the number of times each mass shift was detected. Together, these findings suggest that holobionts differ more in their patterns of molecular abundance changes (akin to gene expression) than in the diversity of mass shifts they can carry out (akin to genomic diversity).

Mass shifts associated with holobionts reflect differences in molecular diversity: By focusing on the differences in mass shift profiles between related molecules, MeMSChem profiling expands metabolomic analysis beyond

288 molecular matches in reference libraries to systemic insights into holobiont
289 biochemistry. If annotated mass shifts reflect single types or classes of molecular
290 modifications catalyzed by enzymes, then disparate mass shift patterns among
291 holobionts may arise for multiple reasons. Holobionts for which the hosts have
292 large genomic differences, such as stony corals and turf algae, may merely
293 possess different biochemical pathways. Among closely-related holobiont types
294 such as the three stony coral genera, the distinct patterns of molecular
295 relatedness may arise from differential expression of shared genes. Yet, the
296 largest disparity among coral holobionts was found by including the mass shifts
297 of molecules that are unique to each holobiont. This suggests that the mass
298 shifts of holobiont-specific molecules largely generate each coral holobiont's
299 unique biochemical profile despite the high degree of taxonomic relatedness
300 among these corals.

301 The mass shifts that differed among holobiont types included differences
302 putatively assigned to single and double-bonded carbon and oxygen, likely
303 among phospholipids and their derivatives based upon the molecules identified in
304 this dataset previously (12) and in the current analyses. These data show the
305 expected lower saturation state of corals relative to algae (19,20) based on the
306 mass shift of *m/z* 2.016 putatively assigned to H_2. Greater fatty acid saturation
307 flexibility can mitigate the damage of elevated temperatures that lead to
308 bleaching in corals (21), suggesting that corals benefit from a higher degree of
309 saturation flexibility and homeoviscous adaptation potential relative to the non-

corals studied here. While desaturations in coral molecules generate double-bonds between carbons, the shift towards gaining H_2O in coral samples suggest these double bonds may be replaced by hydroxyl groups, either directly or through shifts in the relative abundances of molecules. Hydration of phospholipids can lead to conformational changes that alter membrane surface potential and signaling activity (22), suggesting that the higher abundance of hydroxyl groups in corals reflects systemic changes in cell-cell interactions and cellular signaling pathways.

Applications of MeMSChem profiling: MeMSChem offers a new way to analyze existing LC-MS/MS datasets and provides a novel approach for identifying system-wide changes in small molecules across metabolomes. Here we analyzed a dataset collected from one of the most remote places in the world. Like this dataset, other researchers may have LC-MS/MS datasets that cannot be re-sampled or recreated. Therefore, offering a way to gain novel information *in silico* is an attractive proposition for many working with untargeted metabolomic data.

While MeMSChem was applied here to uncover similarities and differences among types of holobionts, it opens doors to answering many more questions. Rather than comparing known groups, MeMSChem may be used to uncover clusters in seemingly homogenous populations (e.g., individuals of a coral species in a common environment, human patients suffering from the same disease, cohorts of offspring growing in a shared location). Annotated mass shifts

can also be searched for and quantified, which may be particularly useful when looking for a ubiquitous process such as antioxidant activity.

If molecules of interest are identified, the mass shifts around them may be used to detect putative sites of known modifications or novel biochemistries. Annotated and unknown mass shifts will require further verification with targeted analyses, such as spiking samples with authentic standards, networking, and examining the mass shifts associated with these standards. Once putative modifications are identified, genetics and molecular biology approaches can be used to confirm or identify the responsible enzyme(s). Such an approach may be particularly useful for tracking molecular changes in time-series samples, a primary need for clinicians (23). Future applications of MeMSChem may also employ a more precise binning approach, taking into account the smaller relative variance at higher masses, changes in MS accuracy across parent masses, and precursor differences. Through this process, the continued application of MeMSChem and the novel form of data it generates will produce a wealth of information from data-rich untargeted metabolomics datasets.

Conclusions: Untargeted metabolomics continues to grow as a tool to examine the complex physiologies of life on Earth. We applied an approach that analyzes untargeted metabolomic data in a new way, based on the chemical relationships between molecules. An analysis of seven coral reef holobionts demonstrated that the relationships between molecules are diverse and distinct between holobiont types. That different types of holobionts had unique

MeMSChem profiles despite high genomic similarity suggests that each possesses physiological capabilities that are not easily identified through traditional genomic approaches. The distinct molecular repertoires identified in each holobiont, coupled with the wide diversity of holobiont types on coral reefs, offers new insights into why this ecosystem is an abundant source of chemical diversity.

Materials and Methods

LC-MS/MS data collection and molecular networking: Samples of seven types of holobionts (hosts and associated viral and microbial communities) including corals, algae, and a fungal mat were extracted in 70% methanol and analyzed with LC-MS/MS (as described in Quinn et al. 2016 [12]; see the SI Materials and Methods for data acquisition details). Files were submitted for molecular network analysis using the online workflow in GNPS (Figure S1; 5), which compares spectral fragmentation patterns and networks related molecules (Fig. S1). Molecular spectra were also compared against reference libraries of known molecules in GNPS. Details of the networking parameters can be found in the SI Materials and Methods.

Identifying aggregations of mass shifts in network pairs: Across all pairs, the difference in mass between each pair of networked molecular features (referred to as 'network pair mass shifts') was searched for aggregations around precise masses. Criteria for identifying aggregations (i.e., redundancies) were

376 established using the similar masses of CO and C_2H_4 (*m/z* 27.995 and *m/z*
377 28.031, respectively; Figure S6; see the SI Materials and Methods for details).
378 The network pairs involved in aggregations were binned by mass shift and
379 counted per sample ('Counts' dataset in the SI 'Processed Data' file). All
380 molecular features involved in redundant mass shifts were then quantified using
381 the Optimus workflow (https://github.com/MolecularCartography/Optimus).
382 Optimus was used to quantify features involved in redundant mass shifts that
383 were present in all files/holobionts, features involved in redundant mass shifts
384 that were present in each holobiont type, and all molecular features including
385 those that were not involved in redundant mass shifts (for normalization of the
386 two former datasets). Molecular abundance data were then used to quantify the
387 molecules exhibiting each mass shift and to quantify the prevailing direction of
388 each mass shift (gaining or losing) in each sample (see the Results and SI
389 Materials and Methods for more details).
390 *Data analysis using Random Forests:* MeMSChem data were analyzed
391 using the ensemble machine learning algorithm Random Forests (16). The seven
392 holobiont types were used as classifiers and MeMSChem data were used as
393 predictors. The out-of-bag error (referred to as 'model error') indicated how well
394 each holobiont type was resolved by the Random Forests model. Permutation
395 tests were used to determine how well the MeMSChem data differentiated the
396 seven holobiont types. These tests were carried out by comparing the model
397 error of the actual data to a distribution of model errors generated from 1000

398 randomizations of the data (see the SI Materials and Methods for more details).
399 The relative importance of each mass shift in differentiating between holobiont
400 types was determined using the Random Forests mean decrease accuracy score
401 and feature importance score (for each holobiont type).

Supporting Information

Figures S1—S6, Table S1—S2, SI Materials and Methods, R code for the Random Forests permutation test, and the processed data for the four datasets represented in Figure 3 can be found in the Supporting Information.

Author Contributions

A.H. developed the method, analyzed data, prepared figures, and wrote the manuscript, D.P. analyzed data and prepared figures, R.Q. analyzed data, I.P. analyzed data and contributed analytical tools, F.A. analyzed data and contributed coding and statistical approaches, E.R. analyzed data, G.W. and B.B. contributed statistical approaches, M.V. collected samples, T.A contributed tools for the mass spectrometry analysis, P.D. contributed tools for the mass spectrometry analysis and interpretation, F.R. developed the method. All authors discussed the results and commented on the manuscript.

Competing Financial Interests

The authors declare no competing financial interests.

Acknowledgements

This work was supported by NSF Partnerships for International Research and Education (PIRE) Grant (124351; F.L.R.) and the Gordon and Betty Moore Foundation (GBMF-3781; F.L.R.). This work was also supported by the NIH

through the grant P41 GM103484 and the NIH grant on reuse of metabolomics data R03 CA211211. European Union's Horizon 2020 Research and Innovation Programme further supported this work under grant agreement № 634402 (to TA and IP). We thank the Deutsche Forschungsgemeinschaft for supporting this work with a postdoctoral research fellowship to D.P. with grant number PE 2600/1-1.

Literature Cited

1. Nicholson JK, Lindon JC (2008) Systems biology: metabonomics. *Nature* 455(7216):1054-1056.
2. Cho K, Mahieu NG, Johnson SL, Patti GJ (2014) After the feature presentation: technologies bridging untargeted metabolomics and biology. *Curr Opin Biotechnol* 28:143-148.
3. da Silva RR, Dorrestein PC, Quinn RA (2015) Illuminating the dark matter in metabolomics. *Proc Natl Acad Sci* 112:12549-12550.
4. Pirhaji L, et al. (2016) Revealing disease-associated pathways by network integration of untargeted metabolomics. *Nature Methods* 13:770-776.
5. Wang MX, et al. (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnol* 34:828-837.
6. Heinonen M, Shen HB, Zamboni N, Rousu J (2012) Metabolite identification and molecular fingerprint prediction through machine learning. *Bioinformatics* 28:2333-2341.
7. Allen F, Greiner R, Wishart D (2015) Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. *Metabolomics* 11:98-110.

8. Duhrkop K, Shen HB, Meusel M, Rousu J, Bocker S (2015) Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc Natl Acad Sci* 112:12580-12585.

9. Van der Hooft JJJ, Wandy J, Barrett MP, Burgess KE, Rogers S (2016) Topic modeling for untargeted modification exploration in metabolomics. *Proc Natl Acad Sci* 113:13738–13743.

10. Simmons TL, et al. (2008) Biosynthetic origin of natural products isolated from marine microorganism–invertebrate assemblages. *Proc Natl Acad Sci* 105(12):4587-4594.

11. Rohwer F, Seguritan V, Azam F, Knowlton N (2002) Diversity and distribution of coral-associated bacteria. *Mar Ecol Prog Ser* 243:1-10.

12. Quinn RA, et al. (2016) Metabolomics of reef benthic interactions reveals a bioactive lipid involved in coral defence. *Proc R Soc Lond* 283: 20160469.

13. Pye CR, Bertin MJ, Lokey RS, Gerwick WH, Linington RG (2017) Retrospective analysis of natural products provides insights for future discovery trends. *Proc Natl Acad Sci* 114(22): 5601-5606.

14. Dinsdale EA, et al. (2008) Microbial ecology of four coral atolls in the Northern Line Islands. *PloS one* 3:e1584.

15. Smith JE, et al. (2016). Re-evaluating the health of coral reef communities: baselines and evidence for human impacts across the central Pacific. *Proc R Soc B* 283: 20151985.

16. Eltahawy NA, et al. (2015) Mechanism of action of antiepileptic ceramide from Red Sea soft coral *Sarcophyton auritum*. *Bioorg Med Chem Lett* 25(24):5819-5824.
17. Breiman L (2001) Random forests. *Machine Learning* 45:5-32.
18. Berk RA (2006) An introduction to ensemble methods for data analysis. *Socio Meth Res* 34:263-295.
19. Harland AD, Navarro JC, Davies PS, Fixter LM (1993) Lipids of some Caribbean and Red Sea corals - total lipid, wax esters, triglycerides and fatty acids. *Mar Biol* 117:113-117.
20. Carballeira NM, Sostre A, Ballantine, DL (1999) The fatty acid composition of tropical marine algae of the genus *Halimeda* (Chlorophyta). *Bot Mar* 42:383-387.
21. Tchernov D, et al. (2004) Membrane lipids of symbiotic algae are diagnostic of sensitivity to thermal bleaching in corals. *Proc Natl Acad Sci* 101:13531-13535.
22. Mashaghi A, et al. (2012) Hydration strongly affects the molecular and electronic structure of membrane phospholipids. *J Chem Phys* 136(11):114709.
23. DeBerardinis RJ, Thompson CB (2012) Cellular Metabolism and Disease: What Do Metabolic Outliers Teach Us? *Cell* 148:1132-1144.
24. Bouslimani A, et al. (2015) Molecular cartography of the human skin surface in 3D. *Proc Natl Acad Sci* 112: E2120-E2129.

25. Petras D, et al. (2016) Mass Spectrometry-Based Visualization of Molecules Associated with Human Habitats. *Analyt Chem* 88:10775-10784.

26. Floros DJ, Petras D, Kapono CA, Melnik AV, Ling TJ, Knight R, Dorrestein PC (2017). Mass spectrometry based molecular 3D-cartography of plant metabolites. *Front Plant Sci 8:429*.

27. Liaw A, Wiener M (2002) Classification and Regression by randomForest. *R News* 2:18-22.

28. Archer F (2016) rfPermute: Estimate Permutation p-Values for Random Forest Importance Metrics. R package version 2.1.1. Zenodo. http://doi.org/10.5281/zenodo.60414.

Figure Captions

Figure 1. Data processing and generation based on a simplified molecular network and two redundant mass shifts. (*A*) GNPS used MS/MS fragmentation spectra to elucidate molecular similarities and network similar molecules (i.e. related molecules). (*B*) Redundant mass shifts between related molecules were identified and annotated to known chemical groups when possible. (*C*) Molecular features that differed by a redundant mass shift were quantified based on MS. Data were generated for (*D*) the number of times each redundant mass shift was observed across all networks, (*E*) the summed abundance of all molecules exhibiting each redundant mass shift, and (*F*) the sum of the differences in abundances between the more massive and less massive molecules for all pairs of molecules connected by a mass shift.

Figure 2. Molecular network of the coral MS/MS dataset. (*A*) The global molecular networks of MS/MS spectra from the coral reef holobiont metabolomic dataset. Each node represents a unique or set of identical spectra (i.e., molecular feature). Lines connecting the nodes represent their spectral similarity, creating subnetworks that can be considered to be molecular families. Circles indicate zoomed-in regions of selected subnetworks shown in (*B*). Node labels represent parent masses and the line labels between the connected nodes represents the mass shift between related molecular features. Nodes with diamond shapes had a spectrum library match, (e.g., Lyso-PAF, as identified by Quinn et al. 2016; 12)

and are further labeled with the molecular names. The size of the nodes indicates the sample frequency in which the spectra were found.

Figure 3. Results of tests measuring the extent to which holobionts were resolved by MeMSChem profiling. (*A*) A visualization of the first two dimensions of a Random Forests proximity matrix of the number of times that each redundant mass shift was identified (counts data). The color of the filled circle represents the holobiont type of the sample while the color of the halo around each circle corresponds to the holobiont type it was placed in by the Random Forests model (i.e., if the circle and halo are different colors the model incorrectly categorized the sample). (*B*) An analogous representation of (*A*) for the summed abundances of molecules grouped by the mass shifts they exhibit among only the molecular features present in all holobionts. (*C*) An analogous representation of (*A*) using the difference in abundances of molecules 'gaining' minus 'losing' a mass, summed by the mass shift they exhibit among only the molecular features present in all holobionts. (*D*) An analogous representation of (*A*) using the difference in abundances of molecules 'gaining' minus 'losing' a mass, summed by the mass shift they exhibit among all the molecules in the dataset. (*E*) A histogram of the permutation test from randomly generated datasets to determine how well MeMSChem profiling resolves each holobiont type based on the model error. Letters above each line correspond to the model error of the actual data in the figure panel matching that letter. The histograms

reflect the model errors of 1000 permutations of the actual data in which pairs were randomly binned while keeping the original proportions consistent. This was repeated for the data in (*A*—*D*), the distributions for which are shown in order and darkening color of counts, summed abundances, differences in abundances in molecules present in all holobionts, and differences in abundances in the entire molecular dataset.

Figure 4. The annotated mass shifts that best differentiated each holobiont type. (*A*) The annotated mass shifts that best distinguish between coral genera based on the mean decrease accuracy of a supervised Random Forests. (*B*) The annotated mass shifts that best distinguish between the non-coral holobiont types. (*C*) The annotated mass shifts that best distinguish the coral holobionts from the non-coral holobionts.

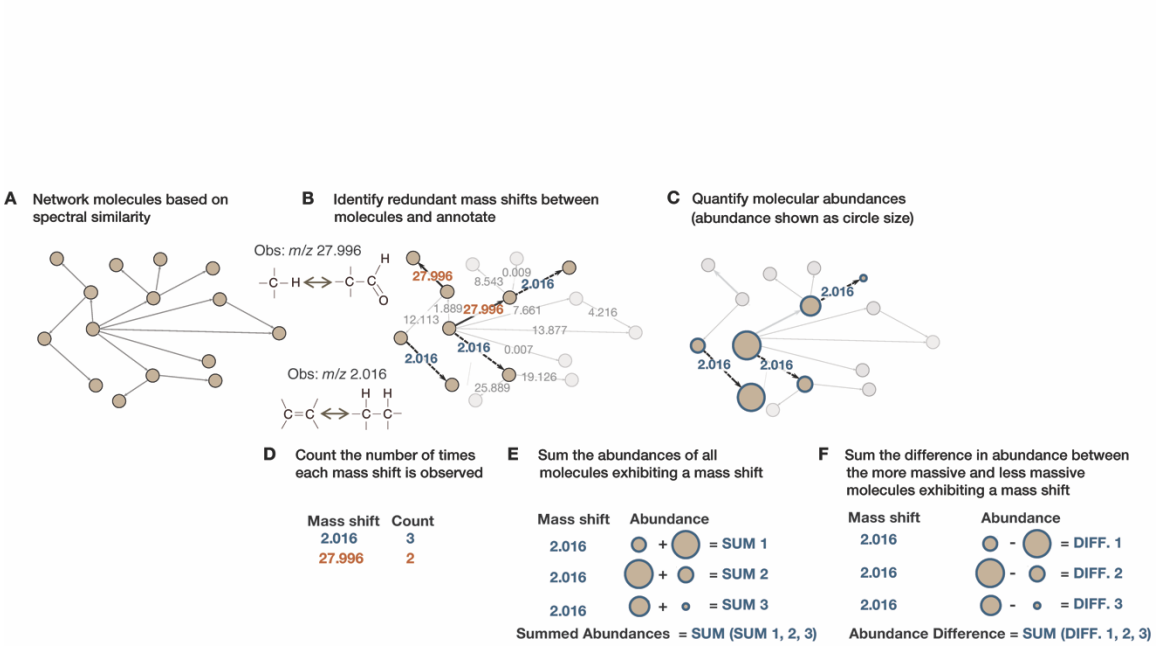

A Network molecules based on spectral similarity
B Identify redundant mass shifts between molecules and annotate
Obs: m/z 27.996
Obs: m/z 2.016
27.996
2.016
C Quantify molecular abundances (abundance shown as circle size)
D Count the number of times each mass shift is observed
Mass shift Count
2.016 3
27.996 2
E Sum the abundances of all molecules exhibiting a mass shift
Mass shift Abundance
2.016 = SUM 1
2.016 = SUM 2
2.016 = SUM 3
Summed Abundances = SUM (SUM 1, 2, 3)
F Sum the difference in abundance between the more massive and less massive molecules exhibiting a mass shift
Mass shift Abundance
2.016 = DIFF. 1
2.016 = DIFF. 2
2.016 = DIFF. 3
Abundance Difference = SUM (DIFF. 1, 2, 3)

Fig. 1

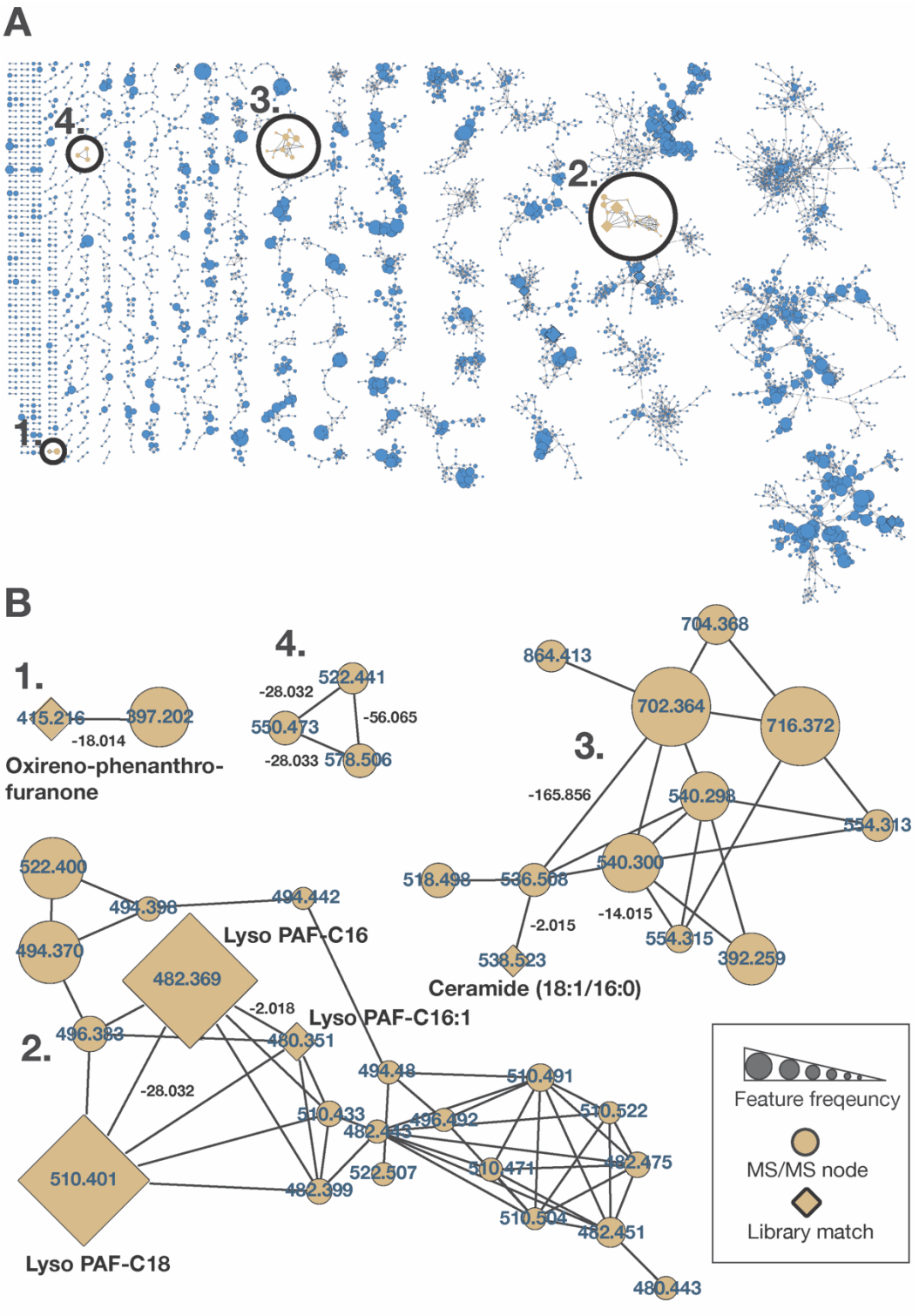
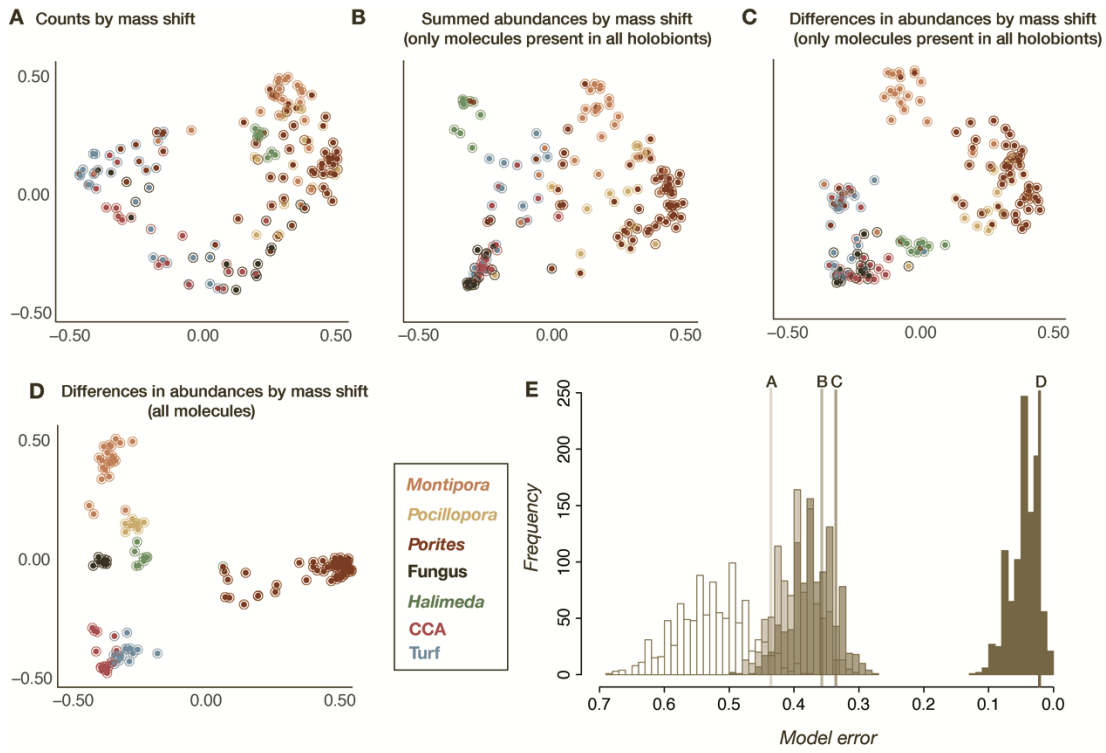
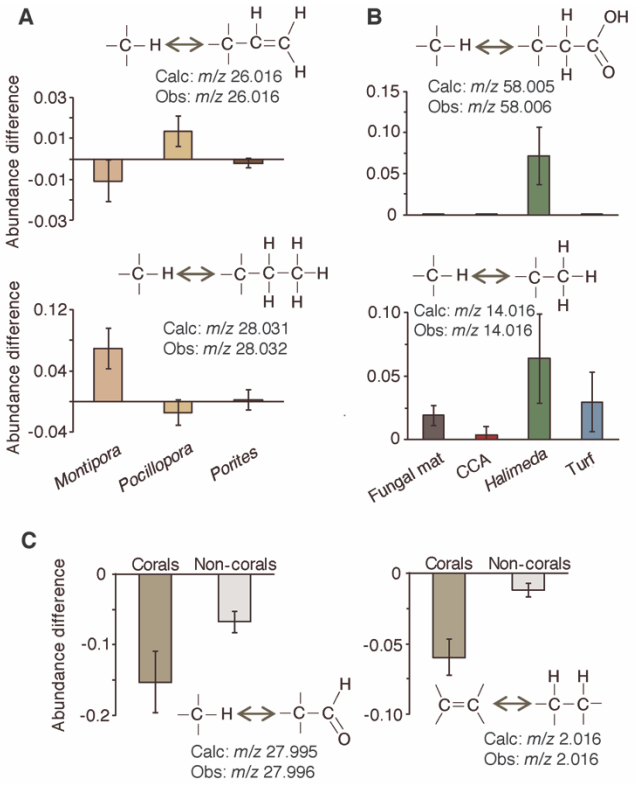

A
1.
2.
3.
4.
B
1.
415.216
397.202
-18.014
Oxireno-phenanthro-furanone
4.
522.441
-28.032
-56.065
550.473
-28.033
578.506
704.368
864.413
702.364
716.372
3.
-165.856
540.298
554.313
540.300
518.498
536.508
-2.015
-14.015
554.315
538.523
392.259
Ceramide (18:1/16:0)
522.400
494.398
494.442
Lyso PAF-C16
494.370
482.369
-2.018
Lyso PAF-C16:1
2.
496.383
480.351
-28.032
494.48
510.491
510.433
496.492
510.522
482.475
510.471
522.507
510.401
482.399
510.504
482.451
480.443
Lyso PAF-C18
Feature freqeuncy
MS/MS node
Library match

Fig. 2



585
586
587 Fig. 3
588
589



590
591
592 Fig. 4